

Dear Recommender,

We thank you and the reviewers for giving us the opportunity to improve our manuscript and submit a revision.

We addressed all issues raised (see below) and modified the main text and supplementary materials accordingly. For ease of reading, reviewers' comments are typed in grey and italicized whereas our answers remained in normal black font. Similarly, we provide a track-change-enabled version of the revised paper.

Best,

Nadia Ponts and Isabelle Fudal
(Corresponding authors)

1-To anonymous reviewer #1's comments:

General comments.

- I do think, that findings should be much more discussed in light of the current state of the art. Authors make a very good job at including relevant information published elsewhere, but in general fail to make a direct connection between the findings of this manuscript and this other work.

We had already added in the discussion mentions to the bibliography regarding the distribution of nucleosomes in other organisms. We have modified this discussion by adding recently released articles.

- Also, the Conclusion section, would benefit from a deeper reflection on the impact of the current results towards our understanding of pathogenicity, since this is mentioned in the abstract and the experiments performed in this study on fungal growth could be used as a baseline to study patterns specific to the infection process. In line with this, I wonder whether the authors could discuss their findings in light of the different lifestyles that each pathogen has and that are highlighted in the introduction.

The objective of this study was to compare plant pathogenic fungi with contrasted genomic structures. To draw conclusions about a link with the lifestyle of these fungi is not possible from this study, *i.e.*, only using genomic and nucleosome organization data. It is true that the abstract and the end of the introduction were confusing and we have modified them accordingly.

- In the Methods section authors present here a new tool to analyse MNase-seq and RNA-seq data. Being such a valuable contribution, it would be useful to detail the highlights of this new bioinformatic tool and how it differs from the ones currently available.

Such details have been included in the Methods and results sections.

- Finally, since this paper will be of interest to both, fungi and nucleosome specialists, I recommend to define the concepts that are field specific (e.g. dispensable chromosomes, nucleosome occupancy (which is not defined until page 6), as opposed to positioning, etc.). We now define the concepts of dispensable chromosomes and nucleosome occupancy vs. positioning.

Other minor comments:

Materials and Methods, Strains and culture conditions. *It would be helpful to introduce this section with a sentence clarifying why the culture conditions were different. If they were performed according to previous studies or current state of the art, etc.*

An introductory sentence has been added.

I would also include in the Materials and Methods sections the analyses that are done with MSTs toolkit. Define how the normalization of the nucleosome signal intensity is achieved, phasograms, etc.

A few additional information on the analyzes performed using MSTs were added in the Methods section to complement the MSTs workflow provided in supp figure 1. Detailed information on the command lines can be found at <https://github.com/nlapalu/MSTs>. If more information is needed, we could provide them as supplementary data.

Results and Discussion, Establishing nucleosome landscapes...: *This is the first and only time that the Ascomycota subdivision is specified. If this has no relevance, it should be removed. Otherwise, the relevance and the subdivision should be mentioned in the Introduction.*

A mention to ascomycetes was added in the introduction section.

Results and Discussion, Genome-wide nucleosome spacing:

- Regarding the definition provided of phasogram "i.e., frequency distributions of read coverage per base genome-wide for all four species". Perhaps the word "read" should be removed, and refer only to the coverage per base genome-wide?

Corrected

- It is not clear to me why the nucleosome signal decays in intensity over these 1,200 bp sliding windows. Would it be possible for authors to clarify that?

Phasograms were initially described in Valouev et al (Nature 2011) as "histograms of distances between mapped reads' start positions aligning in the same orientation". According to this definition, all measurable distances within a given sliding window (*i.e.*, with a given nucleosome taken as a starting point) are considered. In this window, there will be more "short range" distances measured than "long range" ones, because they are more numerous. The wave-like signal will thus be decaying on longer distances.

- At the end of the section authors say that Lmb presenting longer NLR than Lml may be due to the "large AT-rich regions displayed in the genome". Would it be possible for the authors to clarify how the two processes are related?

An explanatory sentence including associated literature has been added to the text.

Results and Discussion, Nucleosome distribution profiles:

- I wonder whether the impact of these findings would be more clear if the differences between read density profiles across fungi were first discussed and then the comparative phasograms for each fungus species would be moved to a new section, where authors explicitly lay out the motivation as to why they are performing these analyses.

We choose the organization of interpreting each nucleosome density profile in a timely manner, *i.e.*, when the density profile itself is presented, to avoid going back and forth for the reader (which can be quite confusing sometimes). The comparisons of these profiles have value only in the light of their respective genome organization. Direct inter-species comparisons without this information strictly provided seems somewhat tricky and to avoid confusions we did not follow that route.

- Also, for B. cinerea, the results are put into context relative to the state of the art, and this could be used to discuss further the results from this study. e.g. What could the increased nucleosome occupancy in BOTY rich regions in fungus grown in axenic culture suggest?

A hypothesis was added regarding the role of a high nucleosome occupancy at un-RIPped BOTY loci in the control of siRNA production during axenic growth.

- I think results would be more clear if they were presented in the same order as in the previous section. Similarly, if in panels B and C in Figures 2-4 the phasograms for the whole genome could be represented (in grey, or shaded) for each species as a sort of reference, that would also be helpful.

We modified Figure 1 to have panels in the same order as Figure 2-4, and change the associated Figure legend and results accordingly. Since the differences with the whole genome are low, they would be difficult to visualize adding grey or shaded representations on the same figure.

- I couldn't help but notice the striking nucleosome density in dispensable chromosome in Lmb. Is this the only dispensable chromosome in this species? I find it astonishing, and I think it would be worth it to discuss this further.

There is indeed only one dispensable chromosome in Lmb. That dispensable chromosome is TE- and AT-rich and, as such share the same nucleosome density as the other TE- and AT-rich regions of the genome.

Results and Discussion, Nucleosome landscapes of fungal gene units:

- Even though this is outside of my area of expertise, I do wonder whether with this dataset it would be possible to somehow perform a comparative analysis of housekeeping genes vs. genes related with pathogenicity to investigate differences in nucleosome occupancy.

In this study, we are using axenic growth conditions, which are conditions where most of the pathogenicity genes are not or only slightly expressed. Comparing pathogenicity genes vs. housekeeping genes is therefore equivalent to comparing expressed vs. not expressed genes *in vitro*, which is already done in the section 'Nucleosome landscapes of gene units according to gene expression'.

- I think this section would also improve if authors discussed the implications of the differences in distances between NDR and ATG sites, as well as provide more details on how they are subject to evolutionary forces.

These differences in distances between NDR and ATG sites are frequently observed. Recent convincing elements published since the initial submission of the paper came in light and are now discussed in the paper.

- Figure 3A seems to be cropped.

It is not, and we are not sure what gave that impression. Nonetheless, Figure 3 was retreated, similar to all other figures, to remove the uneven background that resulted from building the panels. The final rendering is improved.

2-To Ricardo C. Rodríguez de la Vega's comments (reviewer #2)

General comments:

1. *The bioinformatics workflow is hardly described at all, thus leaving out some important details.*

Details have been included in the revised version of the manuscript (see below).

a) in the methods I couldn't find how nucleosome signal intensity was "normalized", but in figures' legends it is stated that these are z-score normalized, is this the only option available in MSTs? Why are there no positive normalized nucleosome signal values on figure 3A?

The information, completed with the details of the tools used for the performed steps, has been included in the methods section. In addition, the word "normalized" may have been misleading, although it is frequently used for z-scoring. Indeed, z-scoring is, in essence, a data scaling method. Thus, for clarity sake, we propose to replace "normalized" by "scaled" throughout. As briefly indicated in the original version of the manuscript's methods, Matlab was also used here. The information has been made explicit in this revised version of the paper.

b) as far as I can tell MSTs reports the standard deviation of inter-nucleosome distances. In the preprint standard errors are reported, for me it is not clear the reason behind, what was the sample size used to transform SD to SE?

There was no particular reason to it, other than the fact that since scaling and plotting was done in Matlab, linear regression was also performed at the same time and metrics (and the *fitml* function reports SE) automatically collected. Sample size was mentioned in figure legends, *i.e.*, four successive peaks, but such details were indeed lacking in the materials and methods section. This has been corrected.

Authors interpret the minute differences in nucleosome packing between SNP rich and SNP poor region as evidence of "increased frequencies of transient nucleosome positioning events in F. graminearum fast evolving polymorphic islands" (page 7, second paragraph), but no statistical test was applied and if the value oscillates 1 bp around the mean, the difference becomes irrelevant.

The reviewer is absolutely right, and our explanation was obviously unclear since we actually tried to convey that the very little differences in phases are not significant and do not support the observed drops in read coverage. Instead, we hypothesize that the occupancy may explain these profiles rather than positioning. We re-phrased the text to avoid the confusion in our explanations.

c) in my reading "conserved genes" would have almost no variance on nucleosome positioning at the first ATG codon. I actually wonder if the variance in +1 nucleosome and preceding NDR could be due to misprediction of starting codons, would expect the most curated gene models would show a lesser effect as a larger number of annotated ATG would indeed be the starting codon) It seems prediction holds (see figures 6 and 7)

We do not understand well the remark of the reviewer. Looking at conserved genes in Figure 7, nucleosomes are well positioned and the signal is clear.

d) it is not clear to me why the nucleosome signal intensity was not normalized in figure 8, neither why there's no estimation of the variance for the nucleosome depleted region (NDR) before the transcription start site/ATG codon. I couldn't find the sample sizes for the different expression categories, but as the average approaches the true mean for larger sample sizes, I would expect the different expression categories would have different accuracies regarding the actual position of the NDR. If low expressions are a minority, this could explain the erratic period seen across expression categories in figure 8 and the neat periods in figure 7. As of now, I can't phantom how the neat period after the ATG in figure 7 could be generated by the sum of waves in figure 8.

The data are not scaled here because comparisons are not made between fungi (inter-samples) but within the same samples (genes at different loci in the same genome). Thus, scaling is not needed for comparison and would remove all information from the dataset by equalizing everything. Regarding sample size, indeed this was missing and it has been added in Table S6.

e) inter nucleosome distances are more homogeneous with respect to the transcription start site (TSS) than with respect to the predicted translation start site, could the authors offer an explanation? could it be that the first methionine codon is mispredicted as the translation start site? For instance, comparing figures 8C with 8E and 8D with 8F it is clear that nucleosome phasing in Botrytis and Fusarium is much more homogeneous before and after the TSS than with respect to the predicted start of the sequence, how come? Either, there would be two NDR in these species, one at TSS and one at the start codon, or it is the same NDR and the nucleosome at +1 is either less well fixed or ATG is mispredicted (in this case the distance between the TSS and the ATG would oscillate in a narrow range). This is also relevant for the discussion on whether or not the position of NDR is less strict with respect to the TSS.

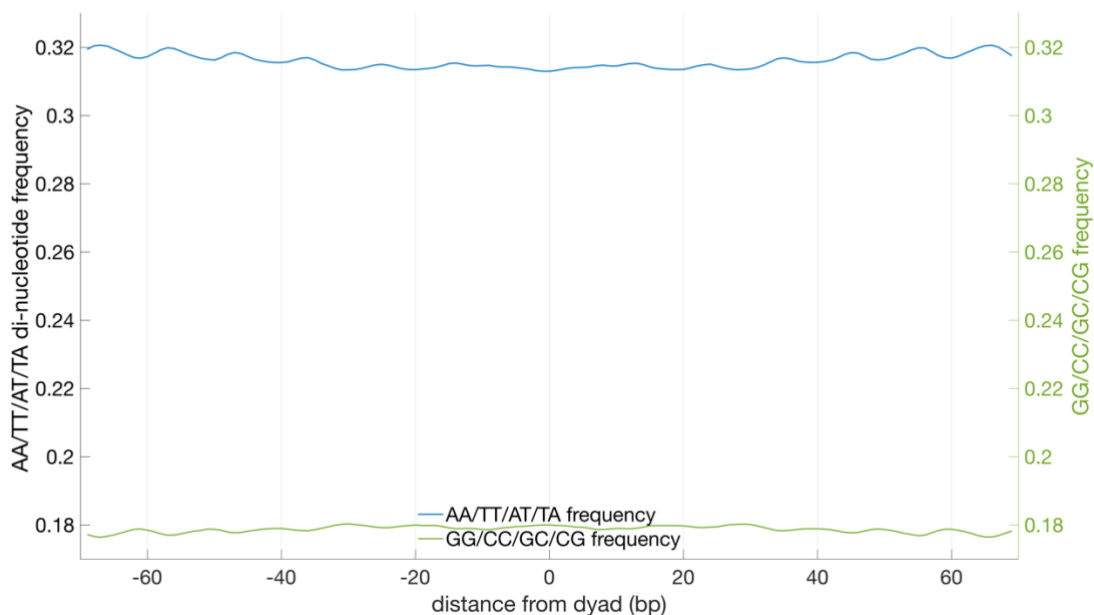
We suppose that the difference in the homogeneity of the phasograms centred to TSS or Translation Start site for *Botrytis* and *Fusarium* is clearly linked to the gene density of fungal genomes and our ability to annotate UTR by bioinformatics. In fact most of adjacent genes with overlapping UTRs have none or partially annotated UTRs. In our case, it is particularly true for *Fusarium* for which only 6110 on the 14143 annotated transcripts have UTRs. Moreover, the length of well annotated genes are longer than short predicted genes such as effectors. So, the combination of these 2 factors, led to a fuzziest detection of the nucleosome profile before and after the Translation Start Site when the whole gene set is used compare to the restricted set with UTR annotation.

2. Some figures appear as a patchwork, with mixed framed and not-framed plots (e.g. figures 3 and 4), clipped ranges (e.g. figure 3C), inconsistent legend and inset position.

We are sorry for the inconvenience. Panel figures were assembled from individual figures that were not all generated in the same format, e.g., tiff or eps, and suffered from some hiccups in subsequent formatting steps. Panel were reassembled for that re-submission trying to avoid file conversion pitfalls.

I also think the choice of presenting dinucleotide frequencies in different scales is misleading (e.g. representation belies that values for dinucleotides involving C and G are lower in three species, in *Botrytis* values ranged from 0.176 to 0.181 for dinucleotides involving C and G but from 0.313 to 0.321 for dinucleotides involving A and T).

Putting both signals on the same scale ends up flattening everything and oscillations become invisible (see below an example of signal with *B. cinerea*). Since this periodicity is what we want to evidence, putting the same y-scale would be counterproductive. By the same token, z-scoring everything to center all values to be represented on exactly the same scale would very much be misleading, since the actual AT/GC content values would be lost. Displaying different y-limits is the best way we found to get an effective visual representation of the oscillating dinucleotide frequencies while preserving the overall AT vs. GC differences. Nonetheless, consistently with the order of panels changed in Figure 1 and the other figures of the paper, we modified the order of the panels of Figure 5.



Why the Rorschach-like plots (figure a5) are not strictly symmetrical?

Dinucleotide profiles were obtained from mapped reads. As we expect that the center of the sequence fragment is the dyad of the nucleosome, at genome scale the profile must be close to be symmetrical, but some differences between both sides could appear due to variability in base content.

Specific points

Please make the life of reviewers easier, mark the line numbers on the manuscript.

We are very sorry about that. We added line numbers to that version.

Could authors comment on whether they expect nucleosome profiles would change depending on the physiological and metabolic state of the sequenced fungi?

We don't want to speculate on that point since we have only data on one growth condition and changes on physiological or metabolic state could lead either to major changes on chromatin accessibility or just lead to local changes, those changes being very transient or on the contrary quite stable, with very different mechanisms possibly underneath these observations.

I think authors overplay the "genome compartmentalization" card, I reckon there are as many phytopathogenic fungi with largely homogeneous genomes (in terms of mobile elements for instance) as there are with compartmentalized genomes (e.g. Torres et al., 2020 doi:10.1016/j.fbr.2020.07.001)

We do not feel that we are overselling compartmentalization of fungal genomes in our paper. On the contrary, we have been careful to choose fungi (all of them being successful fungal pathogens) with various genome structures: homogeneous, weakly or strongly compartmentalized.

I don't understand what authors meant by saying that "nucleosome positioning and occupancies are subjected to evolution" (p1, abstract) and "NDRs positions and intensities are subjected to evolution" (p10, third paragraph). Everything in biology is of course subjected to evolution, but what the authors showed here is that nucleosome phasing varies depending on genome context and species. Whether these genome contexts and species specific patterns are generalizable is not addressed in this preprint.

Elements of discussion relating evolution and chromatin profiles have been developed in the manuscript.

State the distance between the known TSS and the predicted translation start site for Botrytis and Fusarium.

We are not sure to understand what TSS-ATG distance the reviewer is referring to, or what part of the paper is concerned.

What's the average gene size in the genomes? For Fusarium a drop on nucleosome occupancy over the gene body is clear on figure 7C.

Figure 7C show nucleosome signals centered on TSS for *F. graminearum* (in green). We are not sure what "drop" over the gene body the reviewer is referring to, and we assumed here that the deeper valleys at positions ~+700 and +900 are concerned. It is, in our opinion, that speculating on the biological significance of this drop is tricky. Average gene length in *F. graminearum* is 1,749 bp; average exon length is 574 bp (median length 336 bp). Whilst exon length may somewhat suggest the observed drop reveals that feature not visible in *B. cinerea*, caution should be exerted here considering that the number of features annotated with TSS/TTS in *F. graminearum* are very much lower than in *B. cinerea* (see main text); it could just be that signal is less defined.

Minor points and proofreading (suggested changes in upper case)

p1, abstract: spell out MSTs, e.g. "we developed the tool MSTs (FOR MNASE-SEQ TOOL SUITE)" or "we DEVELOPED MNASE-SEQ TOOL SUITE (MSTs)"

Changed

p1, introduction: comparing fungi (an entire Kingdom) with Insects (a class within the kingdom Animalia) is unfair. I think equating number of named species with biodiversity is misleading for a number of factors (size of the community studying them, research effort, inconsistent species delimitations practices across disciplines, etc)

We deleted 'they rank second in terms of species number, behind the insects,'

p2, first paragraph: "important damages in agriculture, human health, and THE environment" ... "and facilitate infection. EFFECTORS CAN BE small proteins"

Changed

p2, first paragraph: Add references to the statement "Upon plant infection, fungi undergo a tightly controlled transcriptional reprogramming..."

Added

p2, first paragraph: how can we speak of "plastic regions" for genomes with "overall large proportion of TE evenly distributed throughout the genome"?

We replaced 'plastic regions' by 'plastic loci'.

p2, first paragraph: please cite the "several recent studies (pointing) out the potential role of chromatin remodelling" instead of the 6-10 years old reviews

There are not many reviews on that topic. We have added a recent review by Collemare and Seidl on the regulation of fungal secondary metabolism.

p2, second paragraph: "hemiascomycetous yeasts. NO comparative genome-wide analyses"

Changed

p2, third paragraph: "four different plant pathogenic fungi ASCOMYCETES showing"

Changed

p3, second paragraph: "micrococcal nuclease digestion of mono-nucleosomes COUPLED WITH HIGH-THROUGHPUT SEQUENCING (MAINE-seq or MNase-seq)"

Changed

p3, second paragraph: not sure to which "contrasted media" authors refer to. Only one media per species was used.

Corrected

p5, second paragraph and elsewhere: homogenize the abbreviation of use of micrococcal nuclease digestion of mono-nucleosomes coupled with high-throughput sequencing

Done

p5, third paragraph: can you comment on whether the estimated linker lengths (14 to 25 bp) are within what's known for other fungi?

These results are commented in the following paragraph comparing with data from other fungi.

*p6, second paragraph: "plants and other higher eukaryotes", really? "higher" to what? We replaced by 'plants, *Caenorhabditis elegans* and humans'*

p7, second paragraph: "regions equally packed with nucleosomes are interspaced with AREAS with lower density"

Corrected

p8, second paragraph: "previously described ~10 bp-periodicities", previously described where?

The previously described 10bp-periodicities refers to the publications cited at the beginning of the paragraph (ref 58-60).

p10, third paragraph: "when the analysis is restricted to CONSERVED fungal genes". Mind what "culture conditions" are not a factor here as they are linked to the species.

The reviewer is totally right. We modified the section.

p21, figure 4 legend: add the reference of Laurent et al. (2017) to explain how "TE and AT-rich regions" were defined

In fact the TE and AT-rich regions refer to Lmb and not to *Fusarium*. A reference to Lmb genome annotation was added (Rouxel et al., 2011).